



# Exploiting the Unused Part of the Brain

Deep Learning and Emerging Technology  
For Future Detectors

CPAD, October 9, 2016

Jean-Roch Vlimant, with input from many



# Introduction



Several detector and instrumentation challenges can be cast into a pattern recognition, regression or classification task.

Deep learning has made tremendous progress in the recent decades, thanks to new technique, computation acceleration, but also more data, and more driving applications (social media data, robotics, ...).

We will look at several ways to cast detector problems into deep learning and other technique.

Accelerating technologies are enabling deep learning. The field of cognitive computing and brain inspired hardware is emerging and promising for going beyond moore's law limitations.

Potential for bringing more elaborated computation closer to the detector in the data processing pipeline (readout, trigger, ...)



# Outline



- Deep Learning Achievements
- The Enthusiastic Industry
- Pattern Detection/Recognition
- Accelerating Technology



# Advanced Machine Learning and Deep Learning *(my selection)*

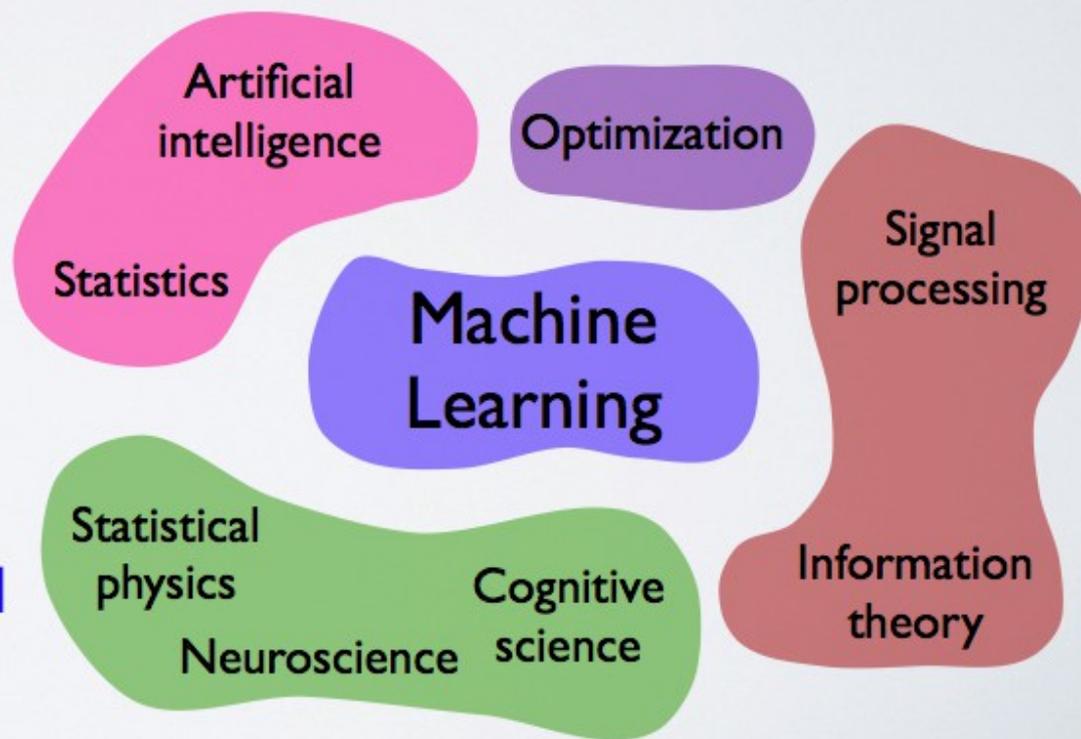


# Machine Learning in a Nutshell



“The science of getting computers to act **without being explicitly programmed**” - Andrew Ng (Stanford/Coursera)

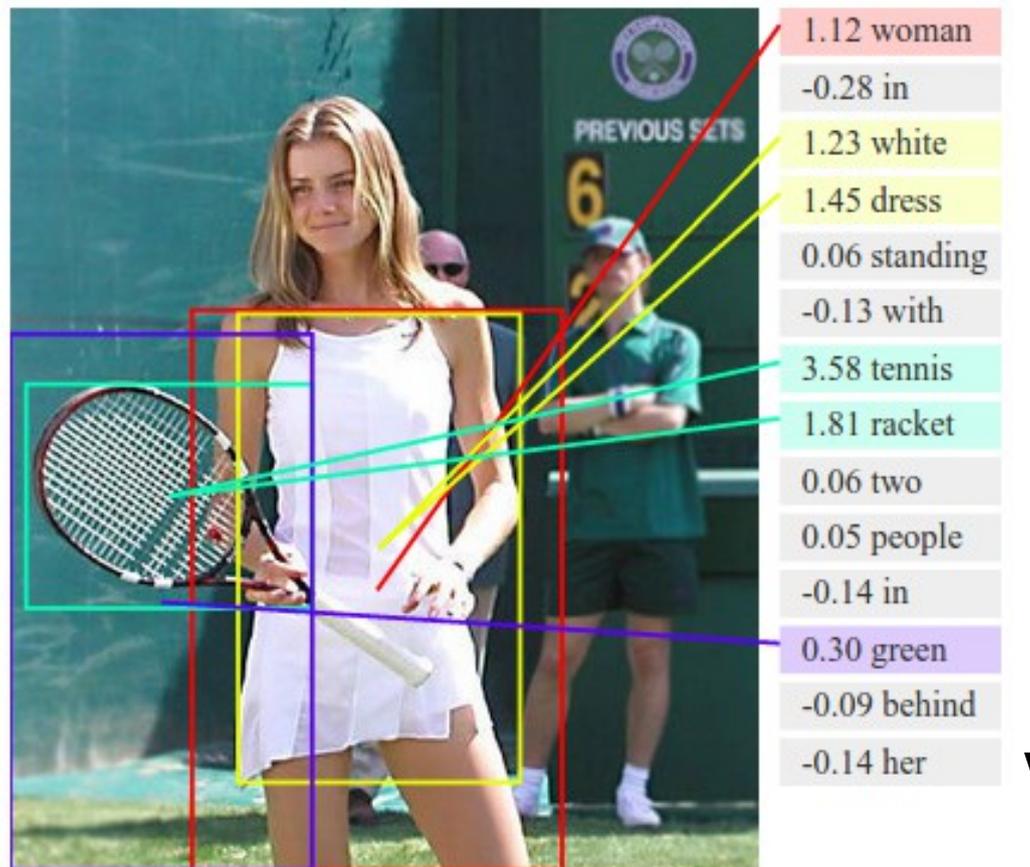
- part of standard **computer science** curriculum since the 90s
- inferring **knowledge** from **data**
- **generalizing** to **unseen** data
- usually **no parametric model** assumptions
- emphasizing the **computational challenges**



Balazs Kegl, CERN 2014



# Scene Labeling



Karpathy, Fei-Fei, CVPR 2015

- Create a description of images
- Generate a decay process description from collision representation, with **application to triggers**



# Scenery Interpretation



Farabet et al. ICML 2012, PAMI 2013

- Group and classify **what each pixel belongs to**
- **Real-time video processing** with deep learning
- Multiple applications to **pileup mitigation, object identification, tracking. All from “raw data”**



# Attention Learning



Near field DL pipeline



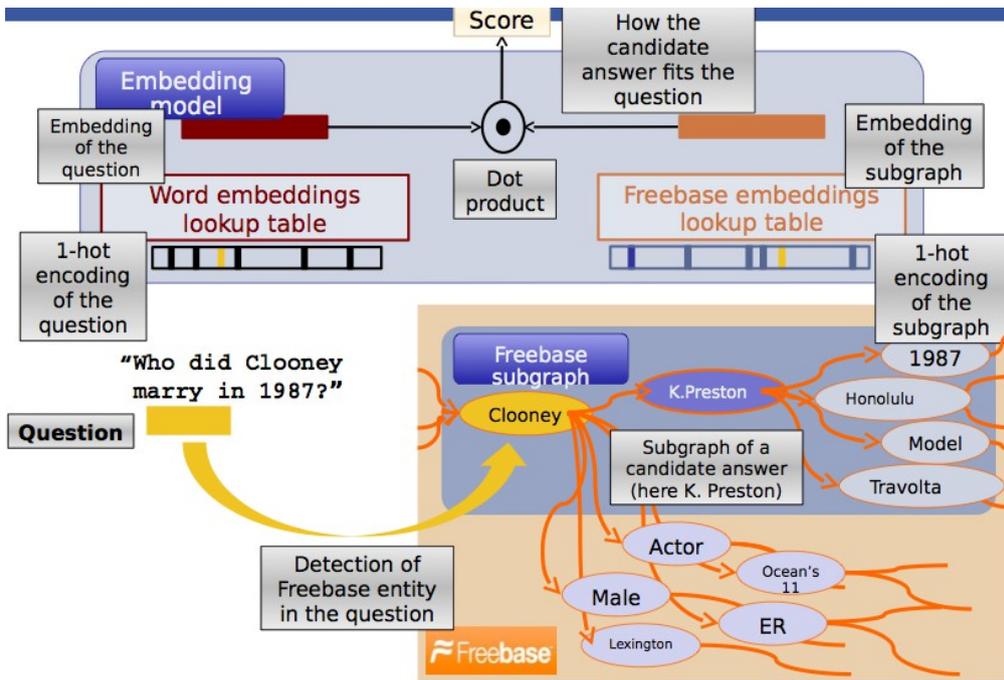
Far field DL pipeline



- Identify people from faces with multiple attention filters
- **Object identification, noise subtraction, ...**

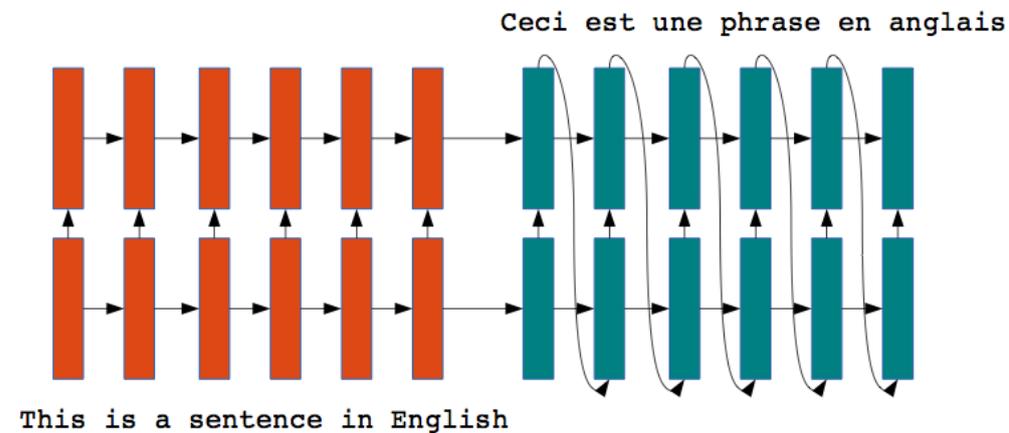


# Text Processing



## [Sutskever et al. NIPS 2014]

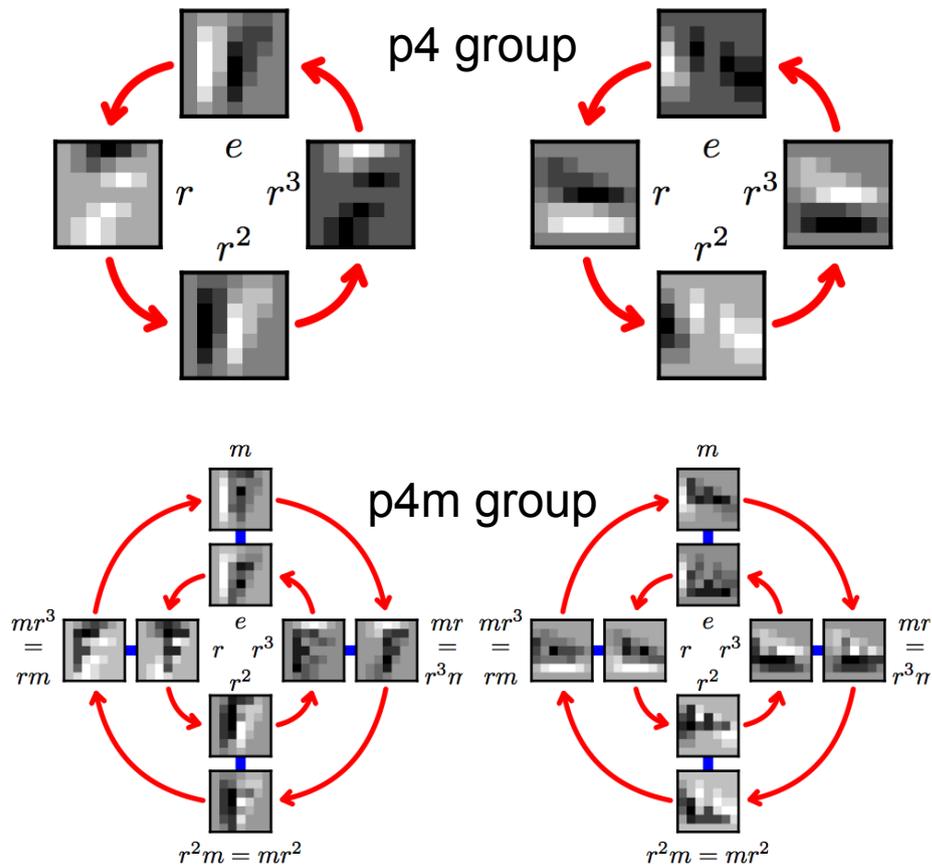
- ▶ Multiple layers of very large LSTM recurrent modules
- ▶ English sentence is read in and encoded
- ▶ French sentence is produced after the end of the English sentence
- ▶ Accuracy is very close to state of the art.



- Question and Answer machine, language translation, semantic arithmetic, ...
- Can the raw data of detector be **interpreted as texts** and **translated into physics descriptions** ?



# Embedded Symmetries

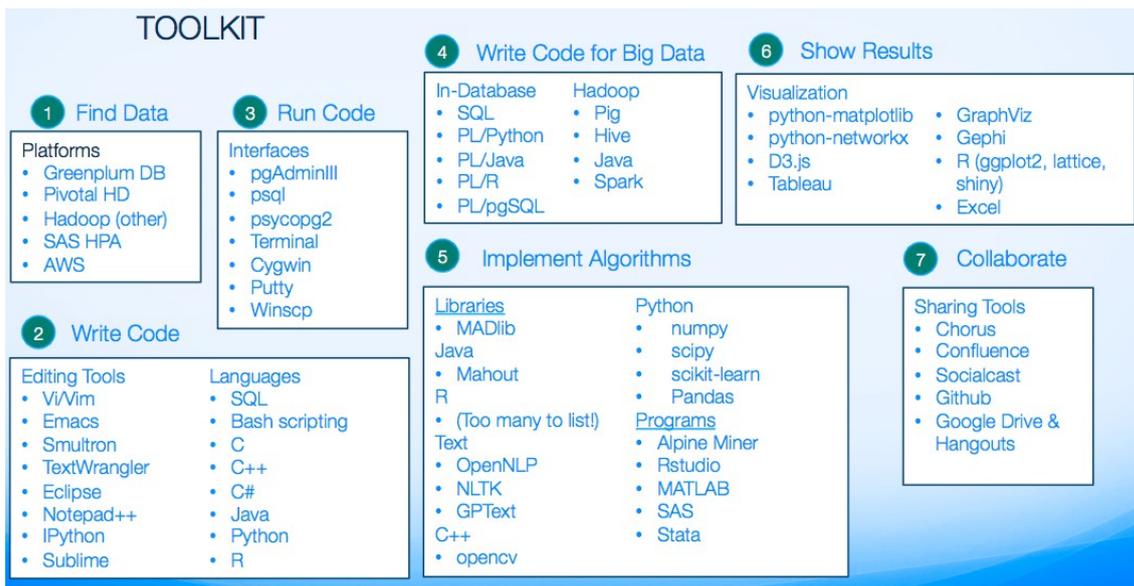


T.S. Cohen, M. Welling ICML2016

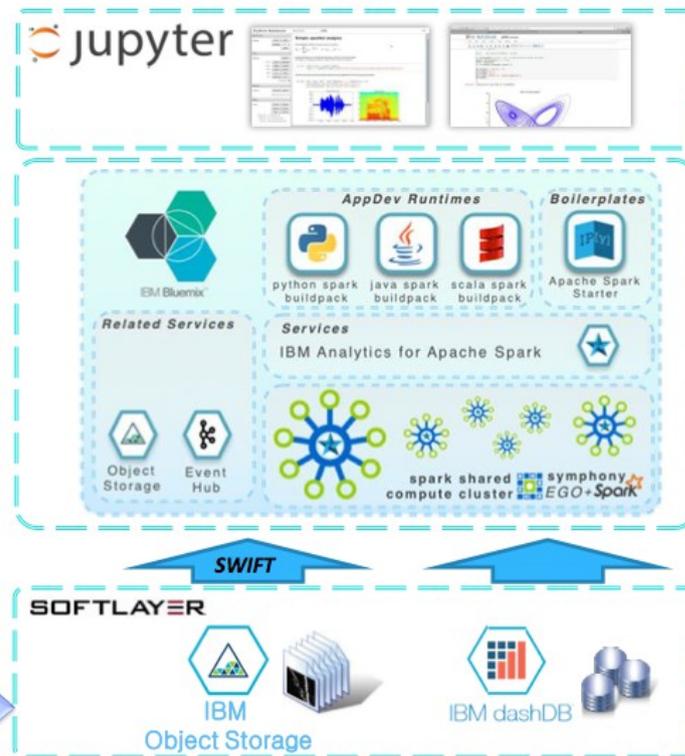
- Introduction of convolutional layers was a **ground-breaking** advancement
- Research on embedding more fundamental symmetries into neural nets
- Symmetries operate on the data or **internal representation** of data
- Next is to implement symmetries of physics to **build physics-specific NN**



# Toolkit and Services



## Spark@SETI



Alex Osterloh @BigDataWizard · 2h

TensorFlow now with HDFS support



**tensorflow/tensorflow**

tensorflow - Computation using data flow graphs for scalable machine learning

github.com

Import of signal data from SETI radio telescope data archives



- Lots of libraries out there, several key components in each major languages. Lots of big-data analytics services offered
- Common theme of going for spark-hdfs support
- Question of having **in-house software** or embracing **external libraries** is very much alive



# Partners in Industry

*(among others, alphas. ordered)*



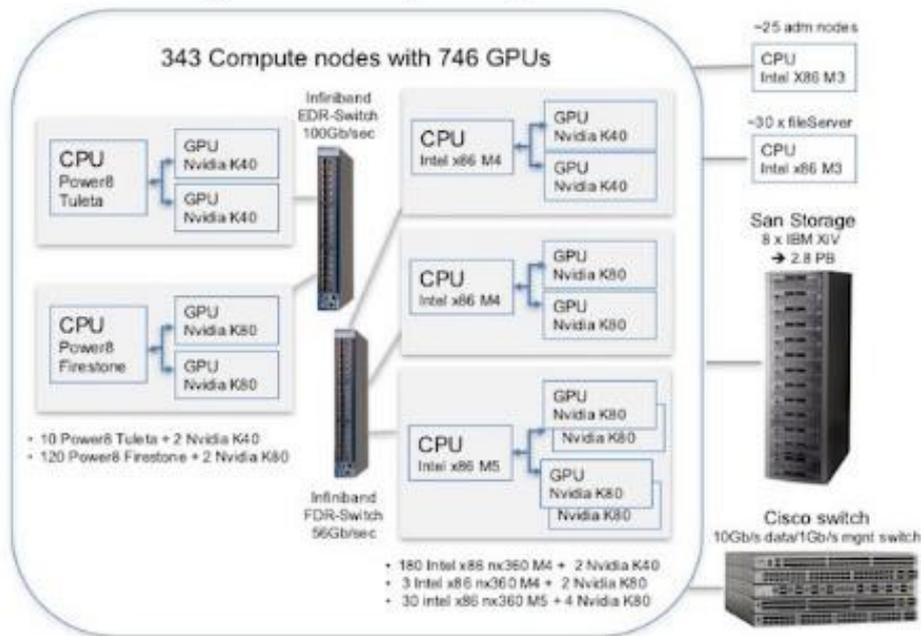
# IBM



## IBM Bluemix Data Analytics Platform



## Cognitive Computing GPU Cluster



- Would participate with providing cognitive computing
- IBM Bluemix opened to development projects
- Enthusiast to work on data quality monitoring and predictivity

<https://indico.cern.ch/event/566167/>

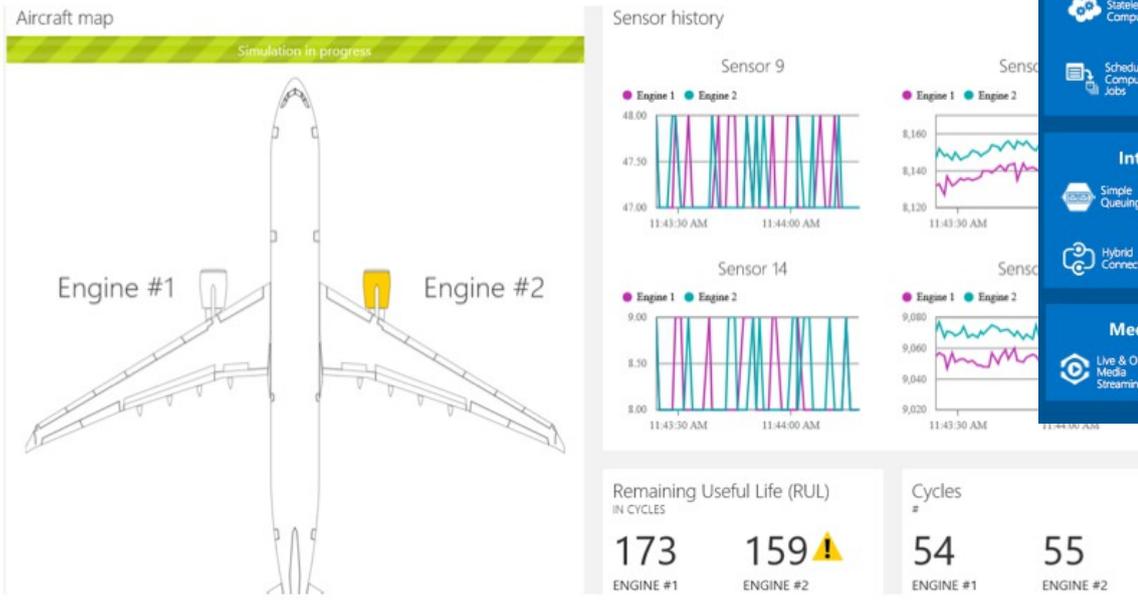


# Microsoft



## Microsoft Azure Services Platform

### Predictive maintenance IoT template



**Platform Services**

- Services Compute**
  - Stateless Compute
  - Distributed Compute
  - Scheduled Compute Jobs
  - Virtual App Streaming
- Web and Mobile**
  - Web Apps Infrastructure
  - API App Infrastructure
  - Mobile Backends
  - Business Process Automation
  - API Management
  - Push Notifications
- Data**
  - Relational SQL Database
  - Data Warehouse
  - Document Database Service
  - Distributed In-Memory Cache
  - Search
  - Simple Key/Value Store
- Integration**
  - Simple Queuing
  - B2B Integration
  - Hybrid Connections
  - Pub/Sub Queuing
- Media & CDN**
  - Live & OD Media Streaming
  - Content Delivery Network (CDN)
- Developer Services**
  - Development Tools
  - Software Development Kits
  - Software Lifecycle Management
  - Application Instrumentation
- Analytics & IoT**
  - Big Data Analytics
  - Predictive Analytics
  - Data Stream Analytics
  - Big Data Storage
  - Data Pipelines
  - Device Data Collection
  - Data Source Management
  - IoT Device Management
  - Mobile Analytics



Familiarity of R algorithms  
+  
Scalability of Hadoop + Spark  
=  
More Accurate Predictions

- Azure platform for big-data analytics
- CNTK Deep Learning Platform
- Looking forward to collaborating

<https://indico.cern.ch/event/514434/>



# NVIDIA



IMAGENET  
Image Classification Object Detection  
**COMPUTER VISION**

Voice Recognition Translation  
**SPEECH AND AUDIO**

Recommendation Engines Sentiment Analysis  
**BEHAVIOR**

Caffe Chainer DL4J Mocha.jl julia K KERAS MarConvNet Microsoft CNTK MINERVA mxnet OpenDeep Purine Pylearn2 TensorFlow theano torch

cuDNN  
**DEEP LEARNING**

cuBLAS cuSPARSE cuFFT  
**MATH LIBRARIES**

NCCL  
**MULTI-GPU**

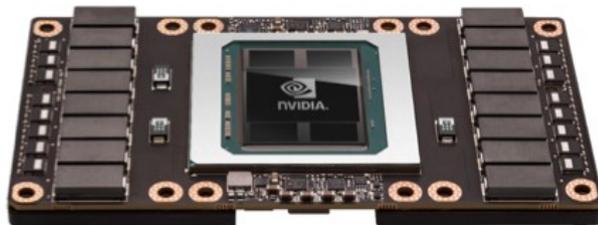
## TESLA P100

Pascal Architecture  
Highest Compute Performance

NVLink  
GPU Interconnect for Maximum Scalability

CoWoS HBM2  
Unifying Compute & Memory in Single Package

Page Migration Engine  
Simple Parallel Programming with Virtually Unlimited Memory



- Major player in GP-GPU Industry
- Supporting drivers and major toolkits
- Actively supporting efforts in adopting deep learning

<https://indico.cern.ch/event/514434/>



# Siemens



- 20 years of experience in machine learning
- Relevant application of re-inforcement learning for wind turbine and machine learning for steel mill optimization
- *“looking forward to continuing the fruitful collaboration with Industrial **Control Systems team**”*



<https://indico.cern.ch/event/514434/>



# Yandex



## Everware. Sharing Research. Reproducible

- › Jupyter-based
- › Docker-empowered
- › github-backed

<http://everware.xyz>

- › Flavours of physics Kaggle challenge  
<https://kaggle.com/c/flavours-of-physics>
- › Machine Learning for HEP summer schools  
<http://bit.ly/mlhep2016>, <http://hse.ru/mlhep2015>
- › Conference on Machine Learning
  - › <https://yandexdataschool.com/conference>
- › Workshop on ML applications in HEP at NIPS'15
  - › <http://yandexdataschool.github.io/aleph2015/>
- › Workshop on Machine Learning in Zurich
  - › <http://indico.cern.ch/event/433556/>

## ML-inspired tools for HEP

- › UGBoost <http://bit.ly/uBoost>
- › GBReweighting <http://bit.ly/GBReweight>

- Active Member of the LHCb Collaboration
- Participated and Organized Outreach in HEP
- “Everware” reproducible research precursor to Swan

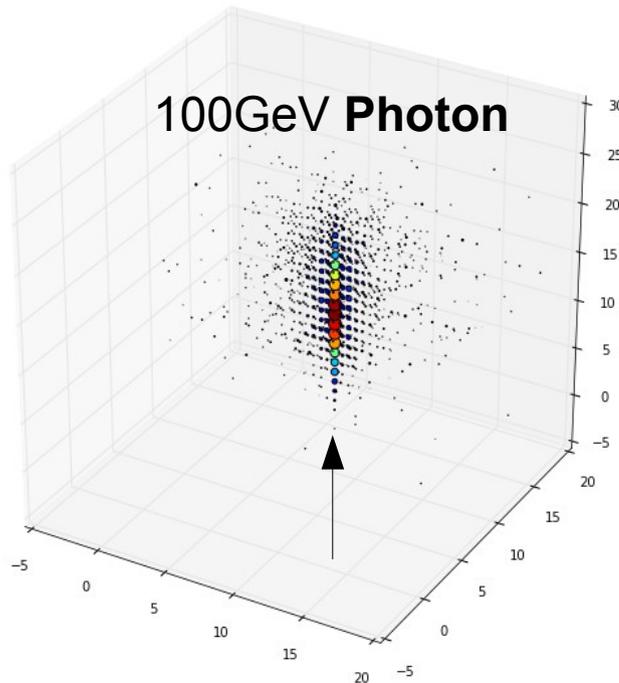
<https://indico.cern.ch/event/514434/>



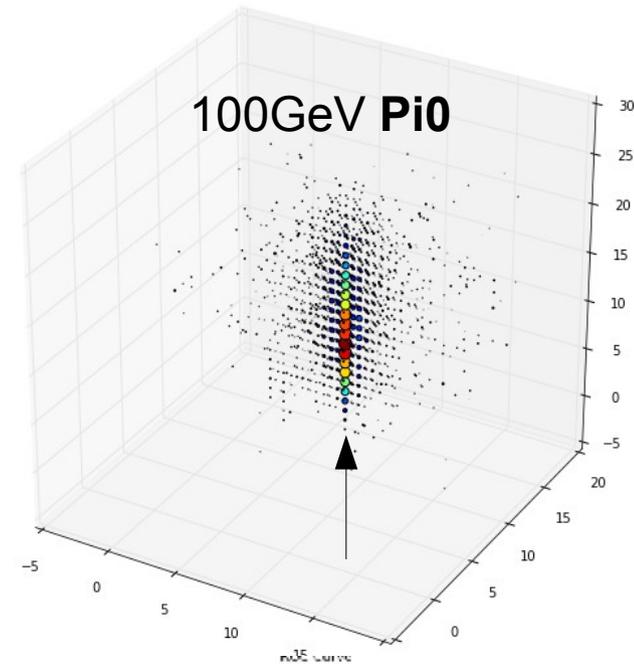
# Application to Intensity and Energy Frontiers *(a selected few)*



# 3D Calorimetry Imaging



≠



LCD Calorimeter configuration

<http://lcd.web.cern.ch>

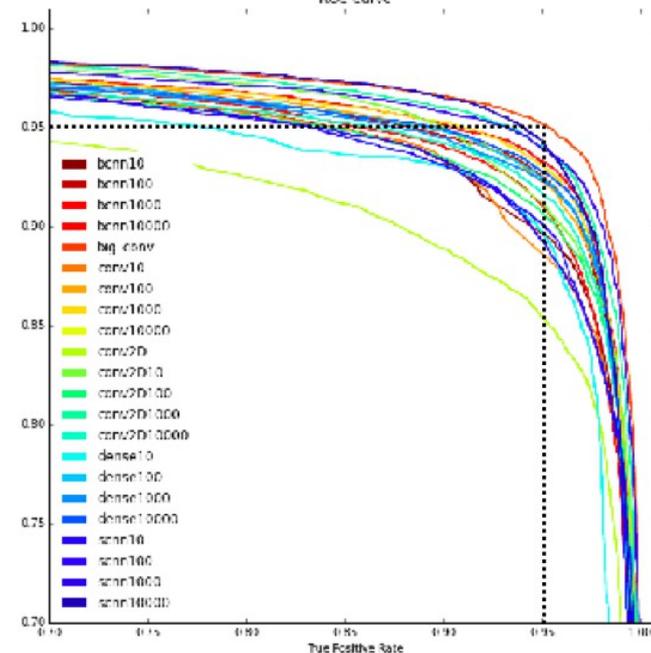
5x5 mm Pixel calorimeter

28 layer deep for Ecal

70 layer deep for Hcal

Photon and pion particle gun

Classification, regression and  
combined models

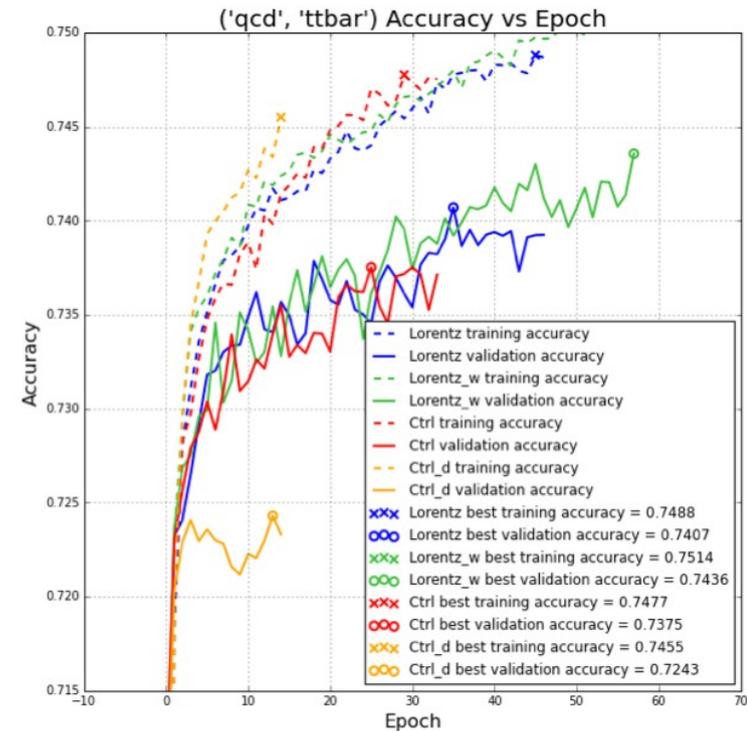




# Collision Event Classification



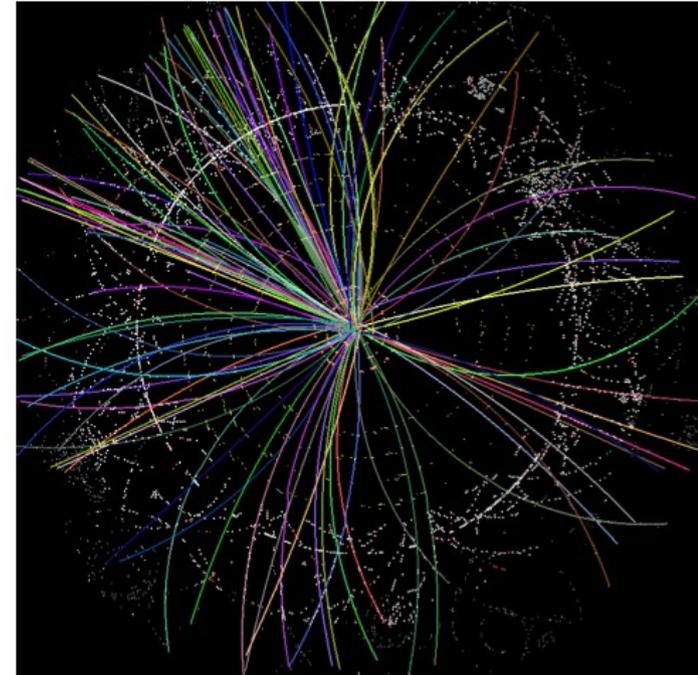
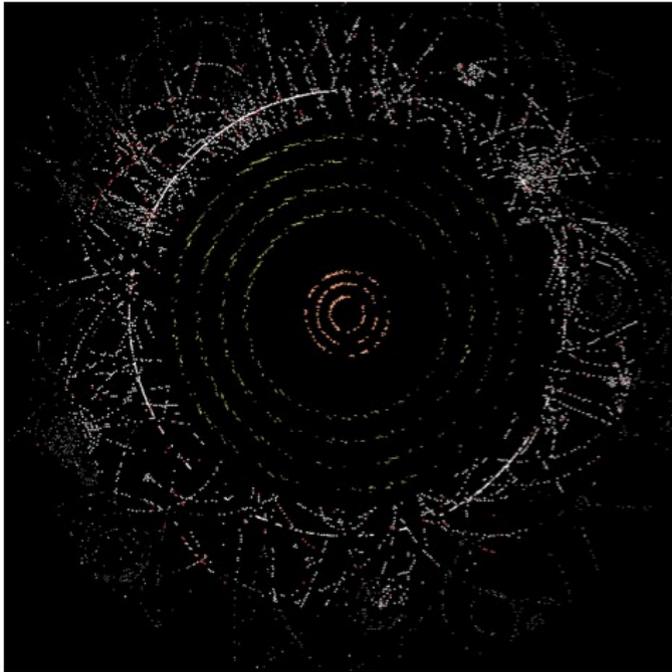
$$\mathbf{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}; v_x, v_y, v_z \in (-1, 1), \quad (1)$$
$$\mathbf{n} = \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad (2)$$
$$\gamma = \frac{1}{\sqrt{1 - \mathbf{v} \cdot \mathbf{v}}} \quad (3)$$
$$B(\mathbf{v}) = \begin{bmatrix} \gamma & -\gamma\beta n_x & -\gamma\beta n_y & -\gamma\beta n_z \\ -\gamma\beta n_x & 1 + (\gamma - 1)n_x^2 & (\gamma - 1)n_x n_y & (\gamma - 1)n_x n_z \\ -\gamma\beta n_y & (\gamma - 1)n_y n_x & 1 + (\gamma - 1)n_y^2 & (\gamma - 1)n_y n_z \\ -\gamma\beta n_z & (\gamma - 1)n_z n_x & (\gamma - 1)n_z n_y & 1 + (\gamma - 1)n_z^2 \end{bmatrix} \quad (4)$$
$$X = \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}, X' = wB(\mathbf{v})X \quad (5)$$



- Full event classification using particle 4-vectors
- Recurrent neural nets, Long short term memory cells
- Dedicated layer with Lorentz boosting
- Step toward event classification with lower level data : low level feature as opposed to analysis level variables



# Charged Particle Tracking



- Perfect example of pattern recognition
- Data sparsity is not common in image processing
- Several angles to tackle the problem. Deep Kalman filter, RNN to learn dynamics, sparse image processing, ...
- Kaggle challenge in preparation

<https://indico.hep.caltech.edu/indico/event/102>

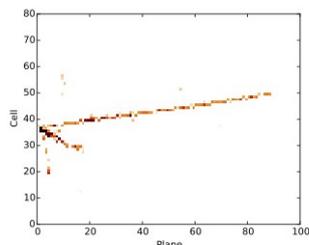
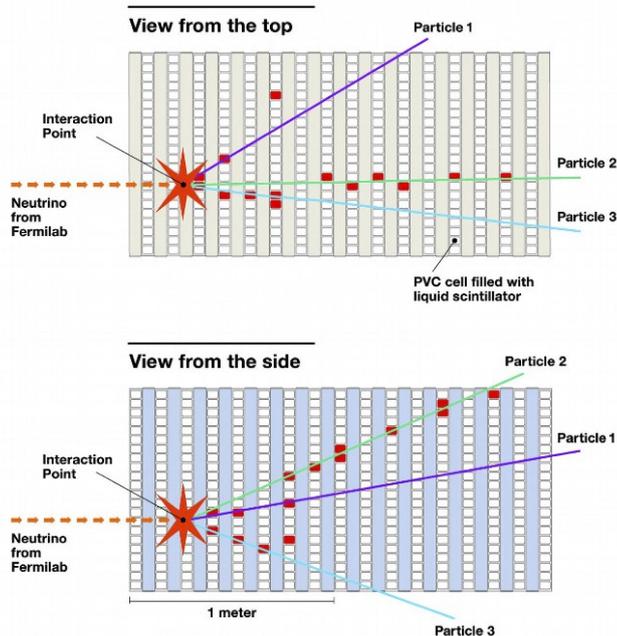
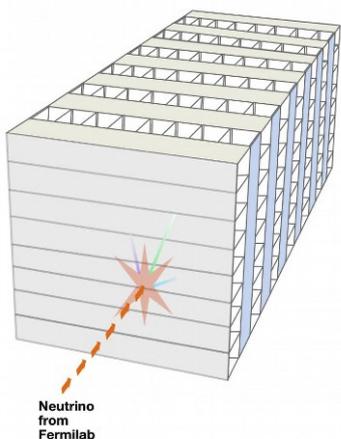


# NOVA Event Classification

Pick something extra from the slides at IBM ?



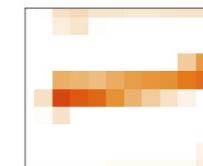
3D schematic of NOvA particle detector



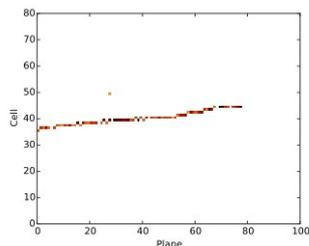
Muon Neutrino DIS CC



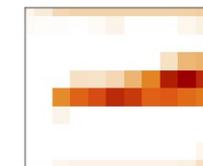
Hadronic Feature Map



Muon Feature Map



Muon Neutrino QE CC

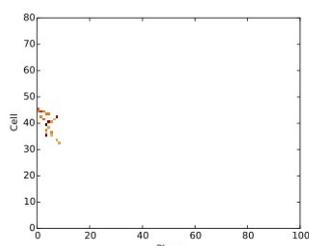


	CVN Selection Value	$\nu_e$ sig	Tot bkg	NC	$\nu_\mu$ CC	Beam $\nu_e$	Signal Efficiency	Purity
Contained Events	-	88.4	509.0	344.8	132.1	32.1	-	14.8%
$s/\sqrt{b}$ opt	0.94	43.4	6.7	2.1	0.4	4.3	49.1%	86.6%
$s/\sqrt{s+b}$ opt	0.72	58.8	18.6	10.3	2.1	6.1	66.4%	76.0%

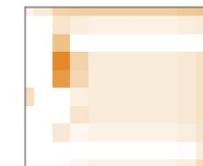
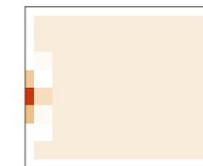
	CVN Selection Value	$\nu_\mu$ sig	Tot bkg	NC	Appeared $\nu_e$	Beam $\nu_e$	Signal Efficiency	Purity
Contained Events	-	355.5	1269.8	1099.7	135.7	34.4	-	21.9%
$s/\sqrt{b}$ opt	0.99	61.8	0.1	0.1	0.0	0.0	17.4%	99.9%
$s/\sqrt{s+b}$ opt	0.45	206.8	7.6	6.8	0.7	0.1	58.2%	96.4%

- 40% Better Electron Efficiency for same background.

<http://arxiv.org/pdf/1604.01444.pdf>



Muon Neutrino NC

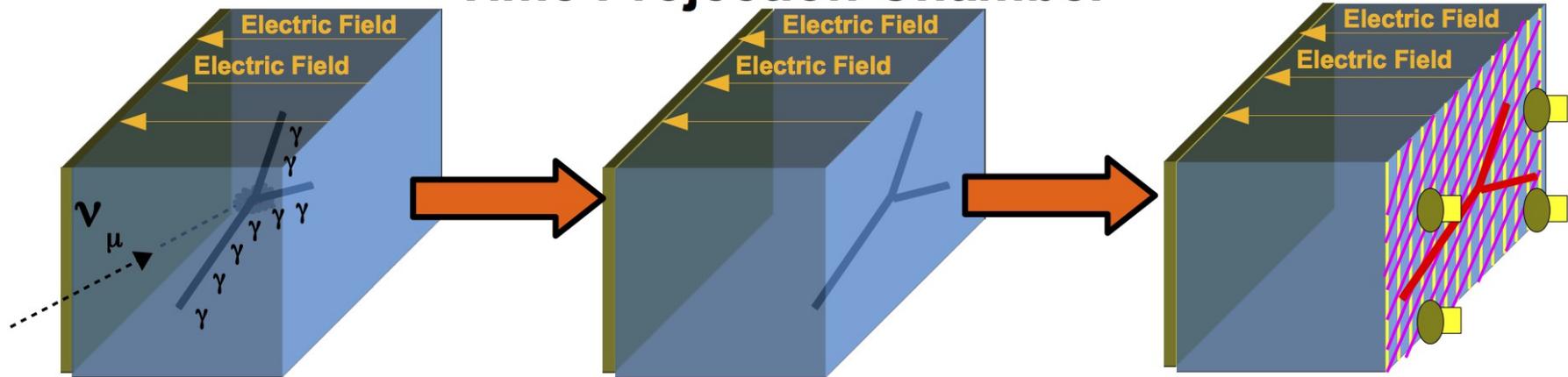




# LarTPC Reconstruction



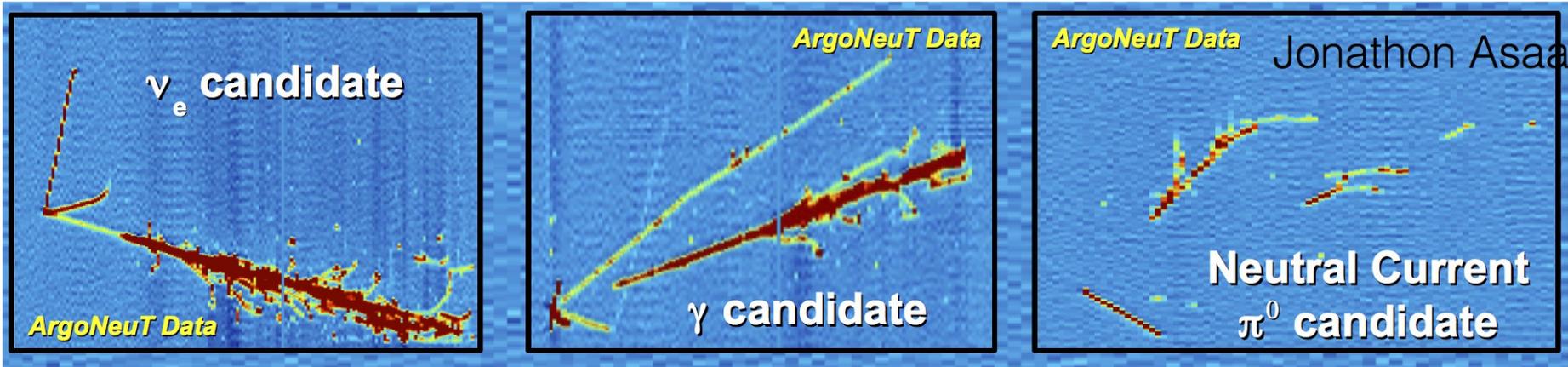
## Time Projection Chamber



Neutrino interaction in LAr produces ionization and scintillation light

Drift the ionization charge in a uniform electric field

Read out charge and light produced using precision wires and PMT's



Tracking, Calorimetry, and Particle ID in same detector.

Goal ~80% Neutrino Efficiency.

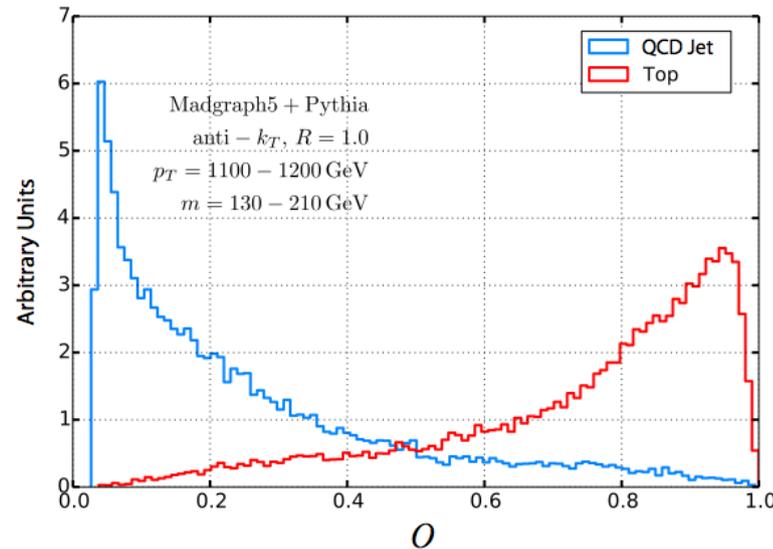
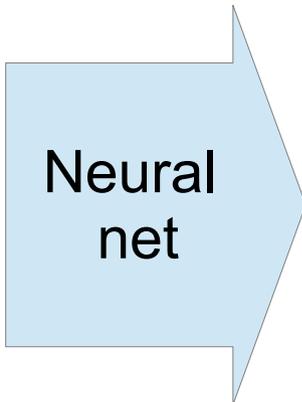
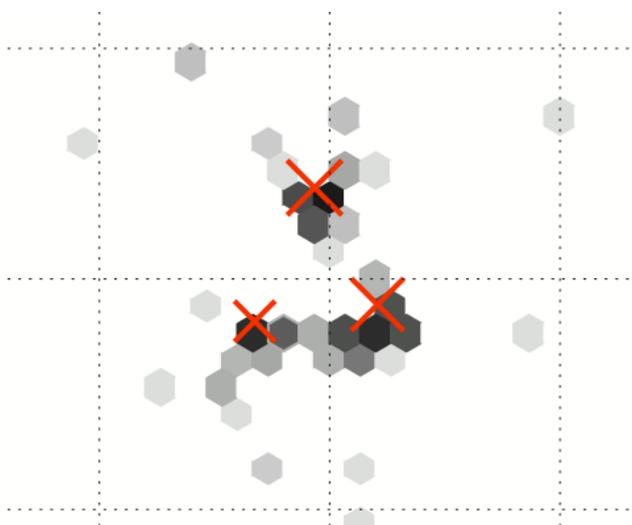
All you need for Physics is neutrino flavor and energy.



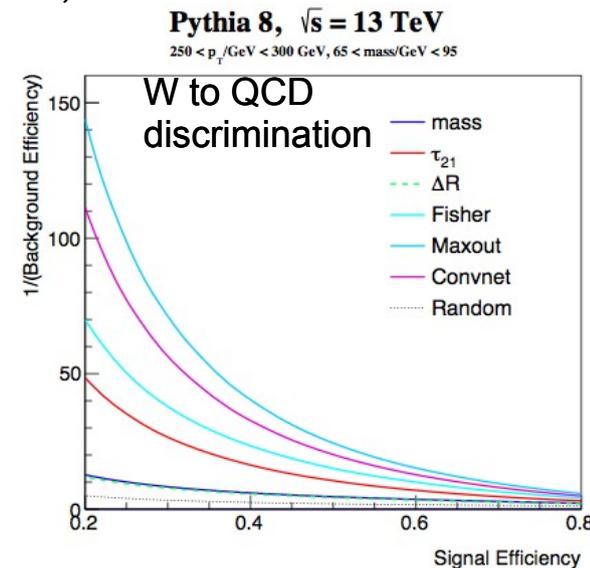
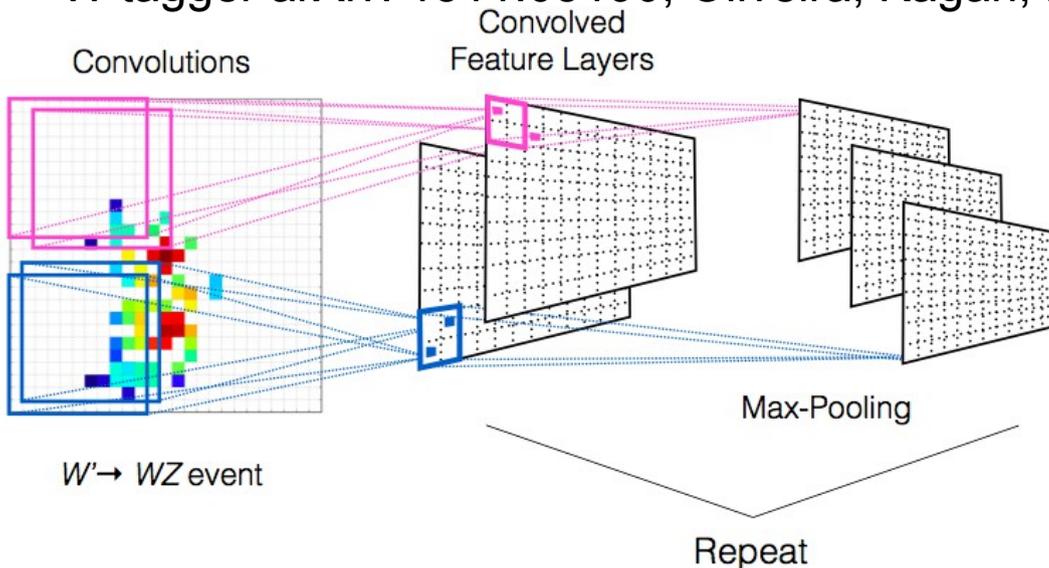
# Particle Jet Identification



Top Tagger arXiv: 1501.05968 Almeida, Backovic, Cliche, Lee, Perelstein



W tagger arXiv: 1511.05190, Oliveira, Kagan, Mackey, Nachman, Schwartzman

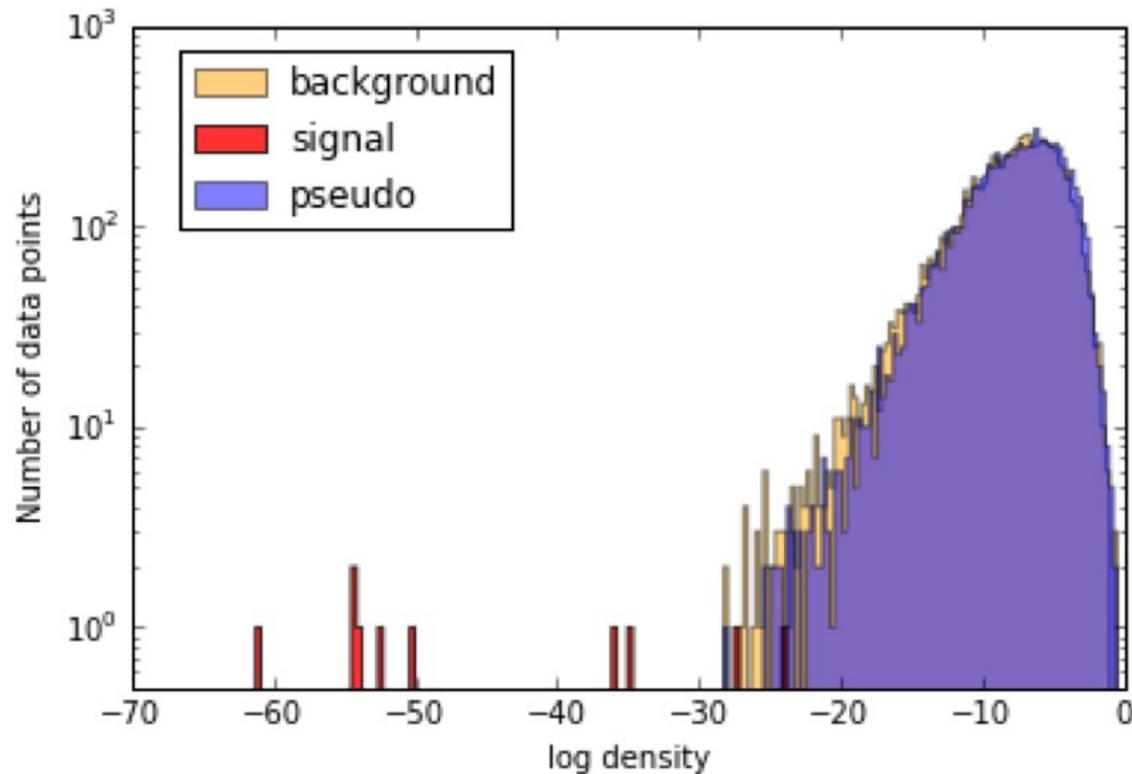




# Outlier Identification



- Train a NADE (arXiv:1306.0186) model on mixture of the known backgrounds
- Use a synthetic dataset with small injected signal
- Log density singles out the injected signal

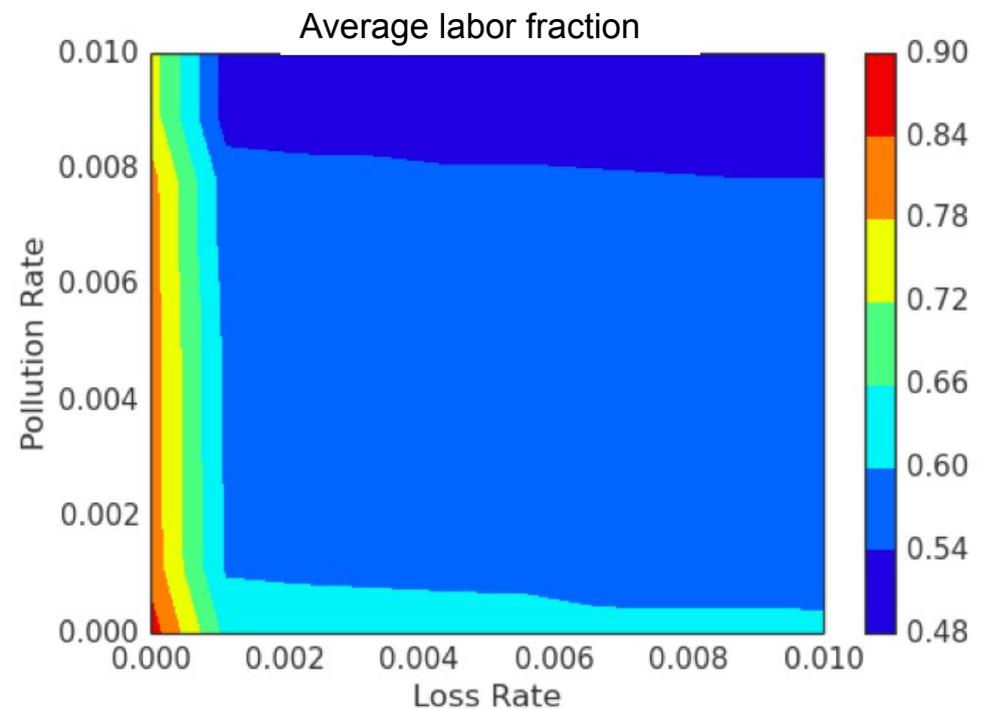




# Anomaly Learning



- Not 100% of the data taken at the experiments are good for analysis (detector effect, calibration, software defect, ...)
- Luminosity block  $\equiv$  23s of beam
- Histograms made per luminosity block are scrutinized by experts to decide on good/bad data
- Several layers of scrutiny, labor intensive
- The machine learning approach
  - Identifies relevant features
  - Calculates percentile per lumiblock
  - Trains rolling classifiers
- **Accepting 1% data loss we could save 40% of the workload on the certification team**





# Cryogenic Anomaly Detection



LHC Logging Service



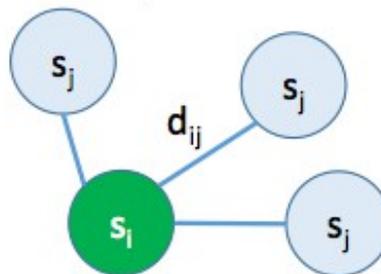
Sensors data extraction



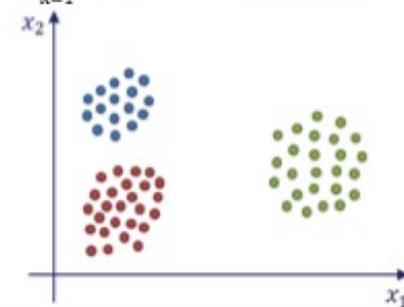
- Building a model based on historical data
- 3 different algorithms
  - Correlation index and KNN-graph
  - K-Mean clustering and probability model
  - Statistics expert-based model

Learning phase

$$E(d_{ij}) = \sum_{j=1}^k d_{ij} * P(j|i)$$

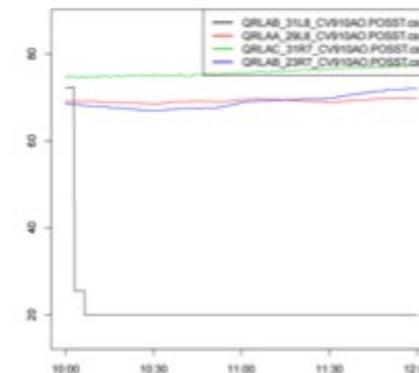
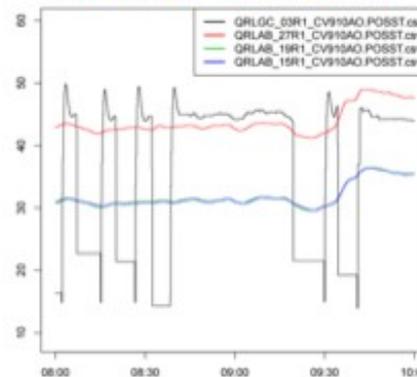


$$\min \frac{1}{C} \sum_{k=1}^c \max_{l \neq k} \left\{ \frac{\sum_i \|x_i - c_k\|}{N_k} + \frac{\sum_i \|x_i - c_l\|}{N_l} \right\} / \|c_k - c_l\|$$



- Use the previous model to detect anomalies
- On-line analysis over a time window of 1 day
- Continuous analysis against thousands of sensors

Anomaly detection



- Project from the LHC cryogenic team

<https://indico.cern.ch/event/514434/>

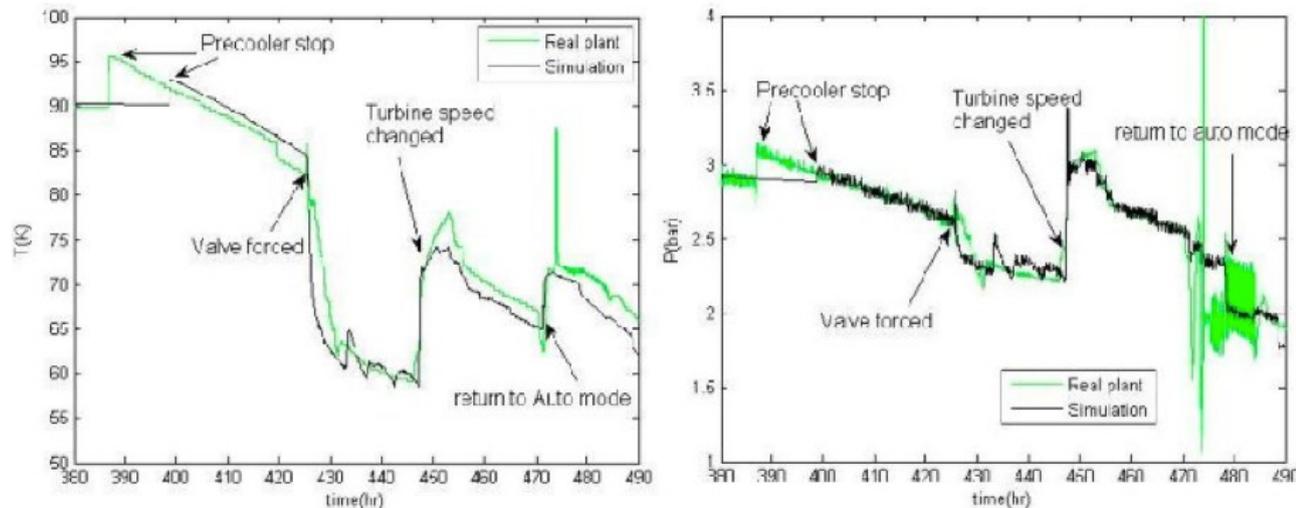


# CMS Magnet Text Book Case



The complete model of the coldbox connected to the Coil Cryogenic System including the superconducting magnet is composed of 3943 algebraic-differential equations. The simulation is performed on a Pentium D 3.4 GHz with 1GB of RAM. In simulation, the complete cooldown of the superconducting magnet from 300K until 5K is performed in 3 days of computation time, hence the simulator ran 7.5 times faster than the real process in average. The simulated cooldown duration is coherent with the observed one (23 days) and the transients of the systems are well simulated.

## Simulation Results :



*Temperature after the first heat exchanger and pressure in the phase separator during manual operations (real plant and simulation)*

- Can defect be detected earlier from sensor data
- Dataset to be shared for collaborative effort
- Technology transfer to monitoring other systems



# Accelerating and Emerging Technologies *(but not restricted to)*



# GP-GPU



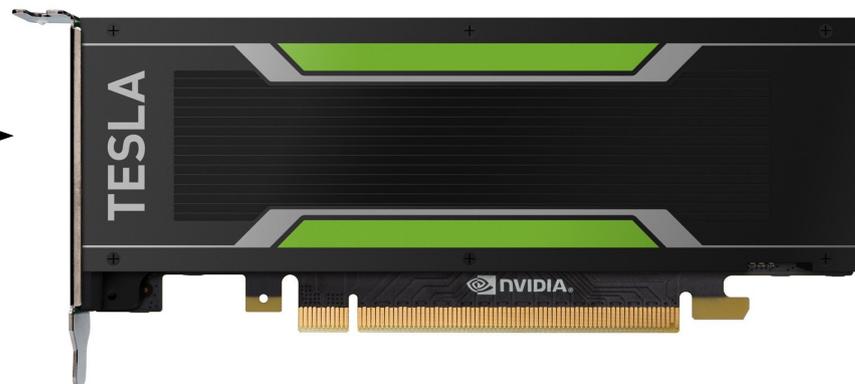
M40

- 7 TFIOps
- 250W



- GPUs are the workhorse for parallel computing
- Enable training large models, with large dataset
- **Deep learning facility clusters**

- Emergence of small GPU
- Not dedicated to training
- Strike the balance between Tflops/\$ for inference
- **Deployment on the grid**

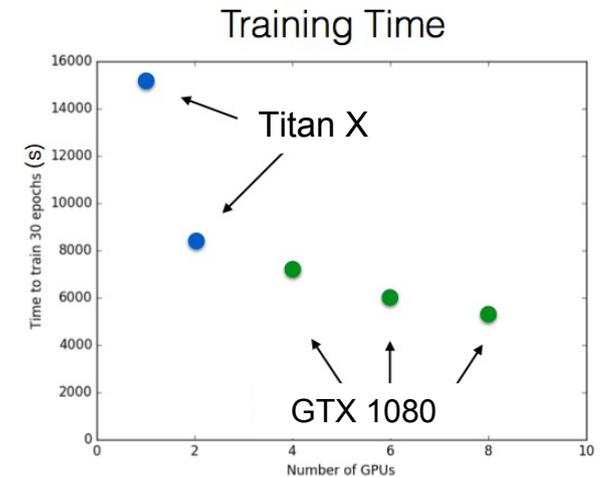
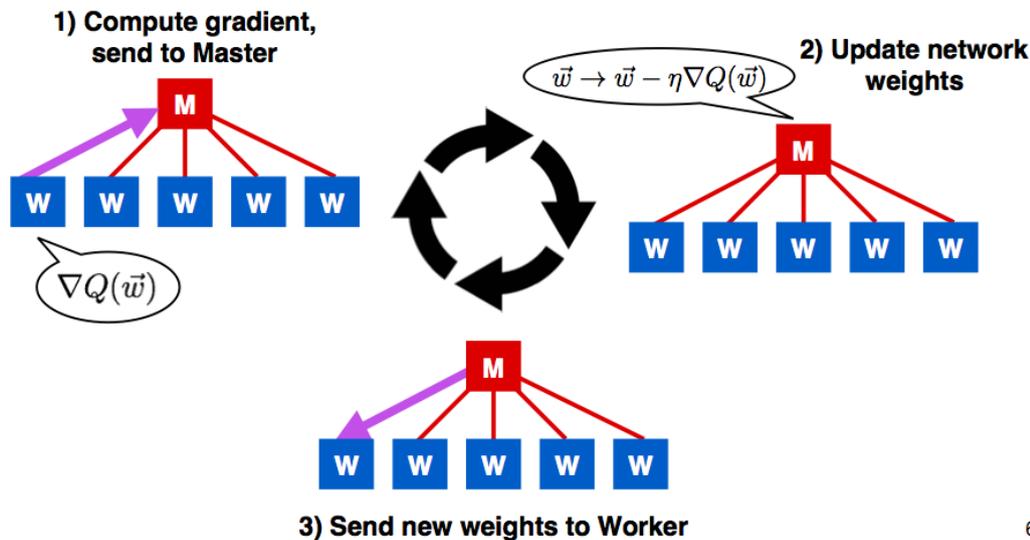


M4

- 2.2 TFIOps
- 50W



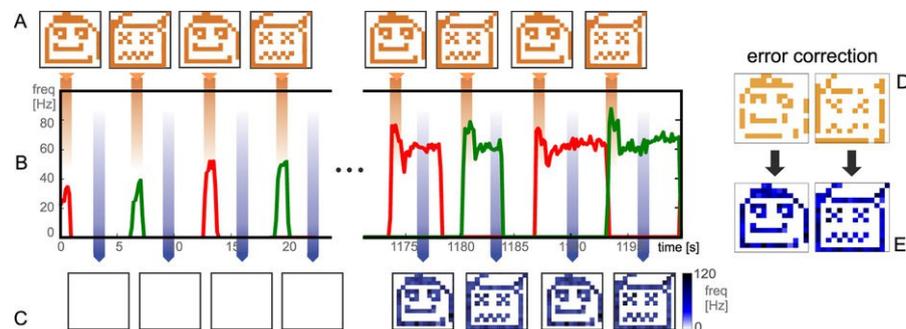
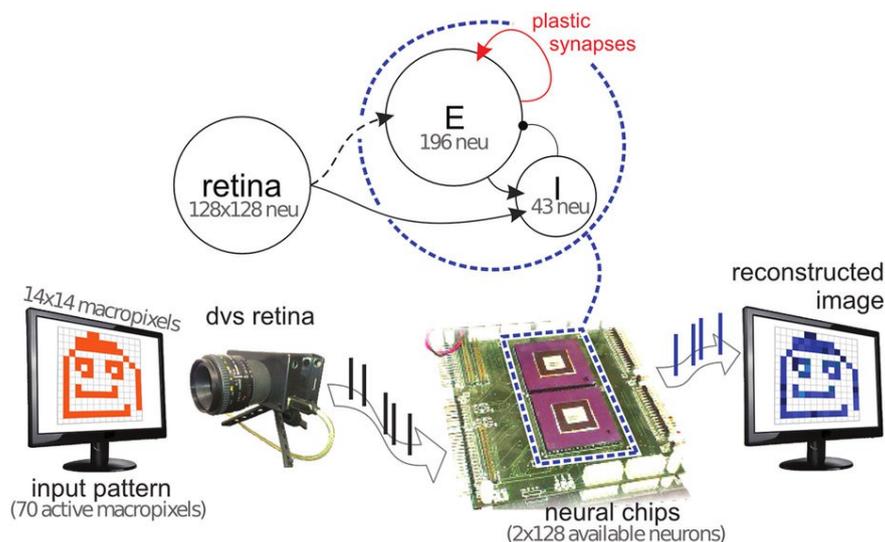
# Distributed Learning



- Deep learning with elastic averaging SGD <https://arxiv.org/abs/1412.6651>
- Revisiting Distributed Synchronous SGD <https://arxiv.org/abs/1604.00981>
- Implementation with Spark and MPI for the Keras framework <https://keras.io/>



# Neuromorphic Hardware

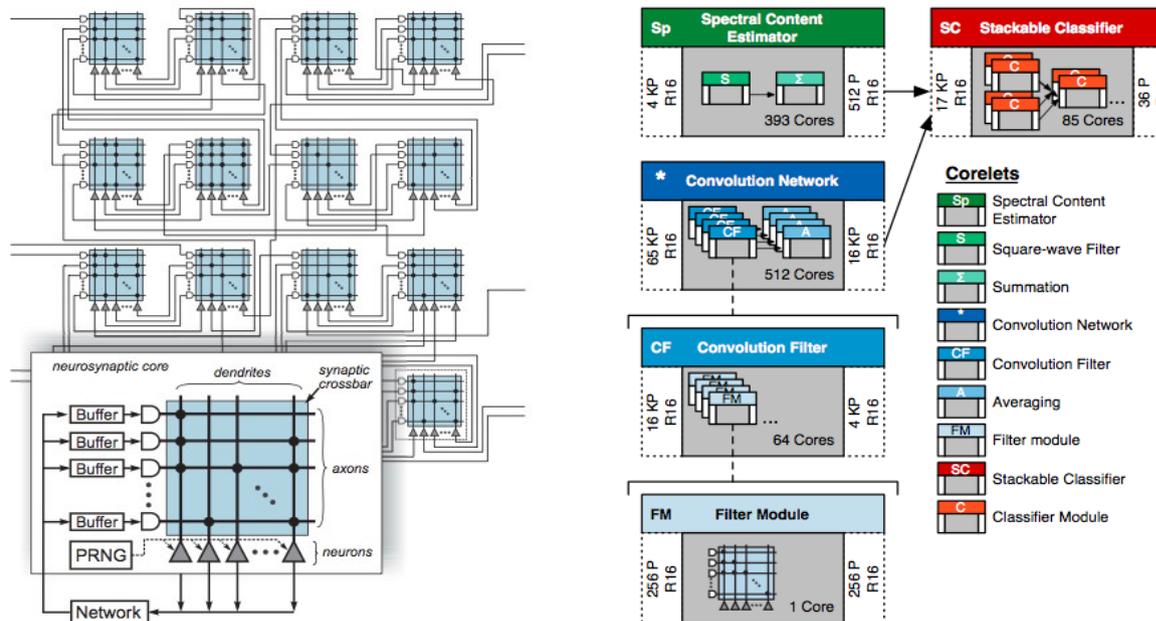


<http://www.nature.com/articles/srep14730>

- Implementing plasticity in hardware
- Process signal from detector and adapt to categories of pattern (unsupervised)
- Post-classified from data analysis or rate throttling
- NCCR consortium assembling to develop this technology further, with our use case in mind



# Cognitive Computing



- Spiking neural net as processing units :  
→ Cognitive Computing Processing Unit : CCPU
- Adopt a **new programming scheme**, translate existing software
- See Rebecca Carney's talk for more details



# Summary



Impressive achievement and promise of modern machine learning and deep learning

From realistic to speculative applicability to field of High Energy and Frontier Physics

Emerging tools and technology to embrace

Partners in industry will take up on our challenges